

國立交通大學

資訊科學與工程研究所

碩士論文

分析樣板比對預測及其交疊區塊動態補償
之應用

Analysis of Template Matching Prediction and Its
Application to Parametric Overlapped Block Motion
Compensation

研究生：王澤瑋

指導教授：彭文孝 教授

中華民國 九十九年 七月

分析樣板比對預測及其交疊區塊動態補償之應用
Analysis of Template Matching Prediction and Its Application to
Parametric Overlapped Block Motion CompensationI

研 究 生：王澤瑋

Student：Tse-Wei Wang

指 導 教 授：彭文孝

Advisor：Wen-Hsiao Peng

國立交通大學
資訊科學與工程研究所
碩 士 論 文



Submitted to Institute of Computer Science and Engineering

College of Computer Science

National Chiao Tung University

in partial Fulfillment of the Requirements

for the Degree of

Master

in

Computer Science

July 2010

Hsinchu, Taiwan, Republic of China

中華民國九十九年七月

分析樣板比對預測及其交疊區塊動態補償之應用

研 究 生：王澤璋

指 導 教 授：彭文孝

國立交通大學資訊科學與工程研究所 碩士班

摘 要

樣板核對預測方法是利用目前編碼的區塊鄰近的點估計其動態向量。此種方法已經被證實有很好的編碼效率，在這篇論文中從理論以及實際統計資料展示了樣版核對的編碼效率，裡面提到了樣版核對的動態向量是跟那樣版區域的重心是有相關性的。樣版核對的效率比 SKIP 預測方法還要好，但是卻比傳統的區域動態補償的預測方式還要來的差，除非在動態向量以及圖片亮度值得亂度很小或亮度值是很高的空間關係性時樣版核對的預測方法跟區域動態補償的編碼效率差不多。我們也示範了如何結合樣板核對和交疊區塊動態補償，以及達到了更好的時間軸上的預測效率。初步的結果與交疊區塊動態補償的比較下，顯示樣板核對和交疊區塊動態補償的結合減少了 MSE 達 2~16%。若與傳統動態補償的預測方法比較時，可以達到更高的 18% 的 MSE 減少量。

Analysis of Template Matching Prediction and Its Application to Parametric Overlapped Block Motion Compensation

Student : Tse-Wei Wang

Advisor : Wen-Hsiao Peng

Institute of Computer Science and Engineering

National Chiao Tung University

ABSTRACT

Template matching prediction (TMP), which estimates the motion for a target block by using its surrounding pixels, has been observed to perform efficiently in inter-frame coding. In this paper, we expose, from a more theoretical viewpoint, the factors that determine the prediction efficiency of TMP. It is shown that the motion estimate found by template matching tends to be the motion associated with the template centroid and that TMP consistently outperforms SKIP prediction, but hardly competes with block motion compensation (BMC) unless both the motion and intensity fields are less random or have high spatial correlation. We also demonstrate how template and block motion estimates can jointly be applied in a parametric overlapped block motion compensation (OBMC) framework to further improve temporal prediction. Preliminary results show that combining TMP with OBMC can yield 2-16% reductions in mean-square prediction error, as compared with the single use of OBMC. The gain is even higher (18%) when the performance is compared with that of the standard BMC.


誌 謝

首先我要感謝指導我快四年的彭文孝老師，在他的指導之下，使我在學術上研究有所成就。老師用嚴謹的態度看待每一件學術的問題，並且對語文與邏輯思考也很講究，因為老師凡事要求因果關係使得我對待一個問題時並不只是看到其表面且會追究其發生原因，以利迅速的解決這個問題。不僅僅在學術上的啟發且在做人處事上面老師使我更為圓融，時時會提點我態度哪邊不妥當，真的很感謝老師的指導與提點，且肯給我機會到世界各地見見世面。在此向我的老師致上無限的敬意。

其次，此篇論文可以完成，我要感謝實驗室的所有同仁們，首先我要感謝陳漪紋學長在我撰寫程式時給我不少的幫助，且不時會跟我討論研究的問題以及想法，引領我的研究步上軌道。感謝實驗室的學弟吳崇豪給我在理論數學上一些提點；感謝學弟李宗霖在我的口試準備、論文撰寫以及實驗上的支持；感謝學長陳俊吉給我在投影片以及英文撰寫的意見；感謝詹家欣學長在這個研究領域的提拔；感謝學妹曾于真替我準備口試的相關事物；也感謝我的同學蔡潤旭、楊復堯以及學弟們陳孟傑、黃家彥陪我度過實驗室歡樂的時光。特別感謝我的室友吳思賢，給我在課業研讀、做人處事以及研究方法上面的意見，並且陪我度過在交大的6年時光。

最後我要感謝我的父母—王宗寶先生、林彩霞女士，沒有他們的栽培與經濟上支持，我無法完成我在碩士的學業、拓展我的世界觀並且專心在研究領域打拼。我也要感謝我的妹妹王柔媚，她給我手足之間的關懷。並且感謝我女朋友林雅婷，他在我研究低潮的時候給我溫暖，研究有所成時給我正面的鼓勵，並且給我心靈上的支持。要謝謝的人時在是太多了，最後還是一句話，謝謝你們，沒有你們就沒有今天的我。

Contents

Contents		ii
List of Tables		iv
List of Figures		v
1 Research Overview		1
1.1 Introduction		1
1.2 Problem Statement		2
1.3 Contributions		2
1.4 Organization of Thesis		3
2 Background		4
2.1 Decoder-Side Motion Vector Derivation		4
2.2 Implementation of DMVD : Template Matching		5
2.3 Overlapped Block Motion Compensation		6
3 Analysis of Template Matching Prediction		7
3.1 Review of Motion and Intensity Models		7

3.2	Error Variance Distribution of TMP	8
3.3	Error Variance Distribution of SKIP Prediction	11
3.4	Comparison of BMC, TMP and SKIP	13
4	Joint Application of OBMC and TMP	15
4.1	Parametric OBMC	15
4.2	Incorporating TMP into OBMC framework	16
4.3	Simulation Results	17
5	Enhanced Inter Prediction with TMP : P^+ Mode	19
5.1	Problem Statement and Theoretical Analysis	19
5.2	Experiment and discussion of P^+ Mode	24
6	Conclusions	29
	Bibliography	31



List of Tables

3.1	Comparison of Mean-Square Prediction Error	13
4.1	Test Conditions	17
5.1	Average Predicted MSE for a 16x16 block	23
5.2	Test Conditions	24
5.3	Average BDBitrate Saving Over H.264/AVC	25
5.4	Bitrate saving and PSNR gain for each sequences and QP setting Over H.264/AVC	27
5.5	Bitrate saving and PSNR gain for each sequences and QP setting Over H.264/AVC	28

List of Figures

2.1	Template Matching Prediction	5
2.2	Rate-distortion curves for DMVD combined H.264 and H.264 [4]	6
3.1	Geometric relation between TMP centroid and block center	9
3.2	Mean-square prediction error surfaces of block \mathcal{B} produced with (a) BMC (b) TMP and (c) SKIP Prediction. The 2nd and the 3rd rows show the error surfaces predicted by Tao and Zheng’s model, respectively. The sequence is Football and the block size used for motion compensation is 16x16.	10
3.3	Inference scheme for SKIP Mode	11
4.1	MSE Reduction: (a) TMP+OBMC vs. OBMC and (b) TMP+OBMC vs. BMC.	17
5.1	16x16 P^+ mode: (a) an example with template centroid located at (1.94,1.94), (b) SMSE surface over different pixel \mathbf{b} , the x and y axis represent the horizontal and vertical indices of pixel \mathbf{b}	21
5.2	Optimal weight surfaces of P^+ Mode produced with (a) Tao (b) Zheng	22
5.3	Predicted MSE surfaces of P^+ Mode produced with (a) Tao (b) Zheng and that of BMC with (c) Tao (d) Zheng	23

5.4 Mode distribution of coding (a) S03 at QP=22, (b) S13 at QP37 25

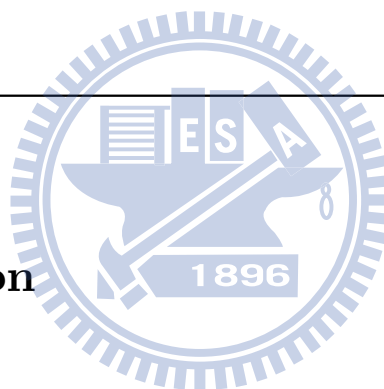
5.5 Mode distribution of proposed scheme and anchor coding (a) S03 at QP=22, (b) S13 at QP37 26



CHAPTER 1

Research Overview

1.1 Introduction



A key problem in video coders using motion-compensated prediction is how to provide a good compromise between the accuracy of motion estimates and the number of bits used to signal them. Variable block-size motion compensation is thus invented, allowing more flexibility in trading off the prediction efficiency and bit-rate. Sometimes a rough representation of motion field is sufficient to provide good temporal prediction, in terms of rate-distortion performance. The decoder-side motion vector derivation (DMVD) is thus proposed to reduce the overhead of motion information by deriving motion at decoder side instead of signaling it. SKIP mode, which is a realization of DMVD, is provided to infer the motion for a macroblock from its neighboring block motion estimates in the state of art H.264/AVC. Somewhat surprisingly, it is often the dominant mode, especially in the low bit-rate coding. DMVD provides a significant performance in a rate-distortion sense, so these techniques should be addressed to further improve the coding efficiency.

1.2 Problem Statement

A decoder-side motion vector derivation scheme has recently been proposed in [4] [7][8][9] to improve the inter-frame coding of H.264/AVC. The approach, often referred to as *template matching prediction* (TMP), finds the predictor for a target block \mathcal{B} by minimizing the prediction error over the pixels in its immediate inverse-L-shaped neighborhood \mathcal{T} (usually termed the *template*). When viewed from motion compensation perspective, it is equivalent to regarding the motion estimate found by template matching as the motion for all pixels in the target block. Since the computation involves only the reconstructed pixels, the decoder can produce the same predictor as the encoder without requiring explicit motion information.

It has been shown in [4] that using TMP in combination with block motion compensation (BMC) can provide 5 - 7% bit-rate reductions over H.264/AVC and the further incorporation of the SKIP mode yields reductions of up to 10%. It is generally believed that TMP can perform very close to BMC when there is a high correlation between the target block and its template. However, it is not appropriated for explaining significant performance of TMP by this intuition. As a result, in this thesis two studies have been made:

1. how to provide theoretical basis for supporting the prediction efficiency of TMP, and
2. why the joint application of TMP and the SKIP mode can give significant bit-rate reductions.

For aiming these issues, two analytical models for intensity and motion field are introduced for providing an in-depth study on the relationships among BMC, TMP and SKIP mode.

1.3 Contributions

Specifically, our main contributions in this work include the following:

- We provide theoretical framework for analyzing mean-square prediction errors(MSEs) of BMC, TMP and SKIP.

- With the help of the MSE model, we have obtained that the template motion tends to be the motion of template *centroid* and TMP can hardly beat BMC unless correlation of intensity and motion field are high.
- We show that TMP consistently outperforms SKIP prediction explaining why the bitrate can be reduced when it is combined with H.264/AVC.
- We find an application of TMP in parametric OBMC. By using the parametric window design, we are able to optimally combine block and template motion estimates to further improve prediction efficiency. Preliminary results show that combining TMP with OBMC can yield 2-16% reductions in MSE, as compared with the single use of OBMC. The gain is even higher (18%) when the performance is compared with that of the standard BMC.
- We develop an enhanced inter prediction, P^+ mode, which achieves bi-prediction performance with only one motion vector. This approach is less complex compared to POBMC and easier to be incorporated into H.264/AVC. It yields about 1-3% BDbitrate saving when it is combined with H.264/AVC.

1.4 Organization of Thesis

The rest of this paper is organized as follows: Chapter 2 contains a review of the Decoder-Side Motion Derivation and Overlapped Block Motion Compensation. Chapter 3 presents a theoretical analysis of TMP, BMC and SKIP prediction. Chapter 4 demonstrates the application of TMP in parametric OBMC. Chapter 5 provides optimal weight and performance of the enhanced inter prediction. Lastly, the thesis is concluded with a summary of our work.

CHAPTER 2

Background

2.1 Decoder-Side Motion Vector Derivation

Decoder side motion vector derivation(DMVD) is a technique that motion information is inferred at decoder side instead of being transmitted by the encoder. This concept is used in the recent video compression standard, such as SKIP prediction and B-direct mode. In both of them motion vectors are inferred with neighboring motion information. Because the inferred motion is inaccurate representation of block motion, the prediction efficiency of DMVD is not as good as that of block based motion compensation. But DMVD still provides a trade-off between bits of motion representation and accuracy of motion vector. This approach,such as SKIP and B-direct, is often the dominant mode when the sequence is compression at low bit rate.

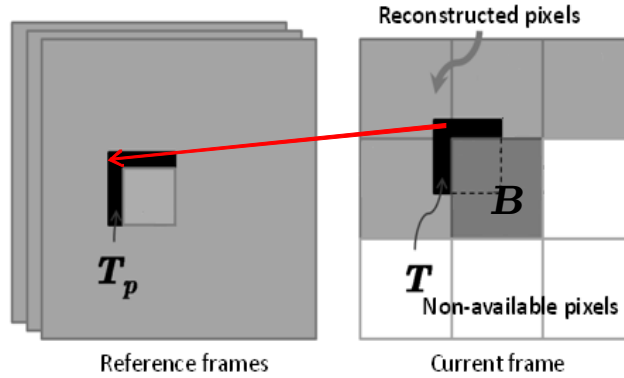


Figure 2.1: Template Matching Prediction

2.2 Implementation of DMVD : Template Matching

Template Matching Prediction (TMP), which uses the correlation of motion and intensity, is one of the realization methods of DMVD. This approach, based on the high correlation between two pixels, was first applied in texture synthesis [11]. In [3], TMP facilitates the synthesis of intra prediction of still image and video coding. Inter prediction using TMP has been studied in [4] [7] [9]. This approach finds the predictor for a target block \mathcal{B} by minimizing the prediction error over the pixels in its immediate inverse-L-shaped neighborhood \mathcal{T} (usually termed the *template*). TMP is further enhanced by averaging multiple candidates to improve and refined the predictor in [9]. It is shown in [4] that TMP provides about 5 - 7% bit rate reduction when it combines with H.264/AVC. Moreover, some approaches of DMVD for Bi-prediction are proposed in HEVC, but in this thesis we only discuss those DMVD employing template matching.

We wish to provide a view of the performance of template matching prediction and the extension implementation for reducing search complexity at decoder side.

As shown in Fig.2.2, the DMVD combined AVC outperforms H.264. Especially in low bit rate, 30% bitrate saving is observed, because that the incorporation of the SKIP mode yields reductions of up to 10% in this sequence.

TMP can improve the coding efficiency, but it also brings some overhead. Because of high complexity of motion search at decoder, the heuristic solution for search is

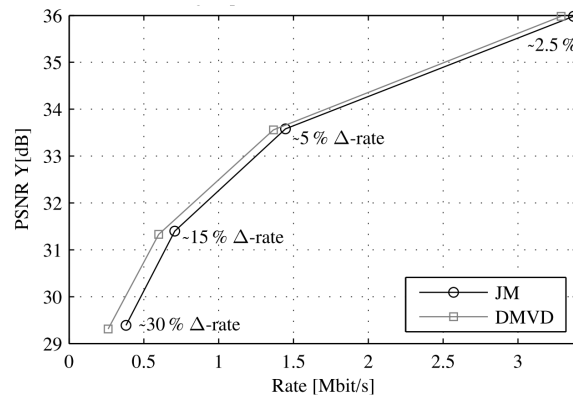


Figure 2.2: Rate-distortion curves for DMVD combined H.264 and H.264 [4]

necessary for TMP. Therefore many approaches such as smaller search range, MVP based predictor and MVP scaling predictor are proposed to reduce the time complexity of DMVD at decoder.

2.3 Overlapped Block Motion Compensation

Overlapped Block Motion Compensation, which is used to combined the TMP in this thesis, improves the motion accuracy for every pixel by considering nearby motion vectors as different plausible hypotheses for its true motion. In order to avoid the blocking artifact raised by block based motion compensation, Overlapped Block Motion Compensation (OBMC) is proposed [5]. OBMC improves the motion accuracy for every pixel by considering nearby motion vectors as different plausible hypotheses for its true motion. In [6], the authors proposed a training based optimal weight by adjusting the ratio of being true motion of each nearby motion vectors. Tao *et al.* [10] reduce the training based weights into a parametric solution with a little coding efficiency lost. In [12], Zheng *et al.* provide an analysis between OBMC and BMC, which explains why OBMC can reduce blocking artifacts. Chen *et al.*[2] also provide a parametric solution for OBMC, by using the motion model in [12], to incorporate the variable block size motion compensation (VBSMC). This parametric solution provides a easier way for finding the optimal weight rather than training the parameters from sequences.

CHAPTER 3

Analysis of Template Matching Prediction

This chapter provides a theoretical analysis to expose the factors that determine the prediction efficiency of TMP. The analysis is carried out based on the statistical models introduced in Section 3.1.

3.1 Review of Motion and Intensity Models

In this section, we review two statistical models used to characterize the motion and intensity fields of video signals. These models will serve as the basis for analyzing the motion compensation error of template matching prediction (TMP) and for determining optimal prediction weights in combining parametric OBMC with TMP.

To analyze the distribution of motion-compensated residuals, Tao *et al.* [10] assumes that the *autocorrelation* function of the intensity and motion fields can be approximated with a quadratic function and an exponential function, respectively:

$$\begin{aligned} E[I_k(\mathbf{s}_1)I_k(\mathbf{s}_2)] &= \sigma_I^2 \left(1 - \frac{\|\mathbf{s}_1 - \mathbf{s}_2\|_2^2}{K}\right) \\ E[v_x(\mathbf{s}_1)v_x(\mathbf{s}_2)] &= E[v_y(\mathbf{s}_1)v_y(\mathbf{s}_2)] = \sigma_m^2 \rho_m^{|\mathbf{s}_1 - \mathbf{s}_2|_1}, \end{aligned} \tag{3.1}$$

where $I_k(\mathbf{s})$ represents the intensity value of pixel $\mathbf{s} = (x(\mathbf{s}), y(\mathbf{s}))^T$ of frame k ; $\mathbf{v}(\mathbf{s}) = (v_x(\mathbf{s}), v_y(\mathbf{s}))^T$ denotes its motion vector; and $\{\sigma_I^2, K\}$ and $\{\sigma_m^2, \rho_m\}$ are their respective variance and correlation coefficient. Likewise, in [12] Zheng *et al.* introduces a motion distribution model assuming that the difference between motion at two pixels obeys the normal distribution:

$$v_x(\mathbf{s}_1) - v_x(\mathbf{s}_2) \text{ or } v_y(\mathbf{s}_1) - v_y(\mathbf{s}_2) \sim N(0, \alpha \|\mathbf{s}_1 - \mathbf{s}_2\|_2^2), \quad (3.2)$$

where α is a constant indicating the degree of motion variation in the horizontal or vertical direction.

Given these models, they both show that the block-based motion estimate tends to be the motion of the block center \mathbf{s}_c , with the mean-square prediction error for pixel \mathbf{s} , $d(\mathbf{s}; \mathbf{v}(\mathbf{s}_c)) \equiv I_k(\mathbf{s}) - I_{k-1}(\mathbf{s} + \mathbf{v}(\mathbf{s}_c))$, given respectively by

$$E[d^2(\mathbf{s}; \mathbf{v}(\mathbf{s}_c))] = \frac{8\sigma_I^2\sigma_m^2}{K} (1 - \rho_m^{|\mathbf{s} - \mathbf{s}_c|_1}) \quad (3.3)$$

and

$$E[d^2(\mathbf{s}; \mathbf{v}(\mathbf{s}_c))] = \epsilon \|\mathbf{s} - \mathbf{s}_c\|_2^2, \quad (3.4)$$

where ϵ is a factor related to the randomness of the motion and intensity fields (the randomness increases with increasing ϵ). According to these equations, the prediction error is larger for boundary pixels, which agrees with the general observation.

3.2 Error Variance Distribution of TMP

We begin by examining the distribution of TMP error variance. To do so requires modeling the template motion estimate. Proceeding as the approach described in [12],

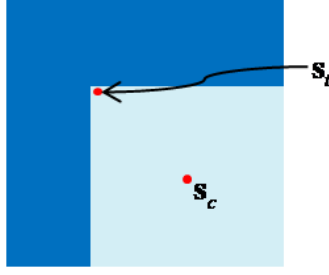


Figure 3.1: Geometric relation between TMP centroid and block center

we can obtain, with the results in (3.3) and (3.4), that

$$\begin{aligned}
 \mathbf{s}_t &= \arg \min_{\mathbf{t}} \sum_{\mathbf{s} \in \mathcal{T}} E[d^2(\mathbf{s}; \mathbf{v}(\mathbf{t}))] \\
 &= \arg \min_{\mathbf{t}} \epsilon \sum_{\mathbf{s} \in \mathcal{T}} \|\mathbf{s} - \mathbf{t}\|_2^2 \\
 &= \arg \min_{\mathbf{t}} \epsilon \sum_{\mathbf{s} \in \mathcal{T}} [(x(\mathbf{s}) - x(\mathbf{t}))^2 + (y(\mathbf{s}) - y(\mathbf{t}))^2] \\
 &= \left(\frac{\sum_{\mathbf{s} \in \mathcal{T}} x(\mathbf{s})}{|\mathcal{T}|}, \frac{\sum_{\mathbf{s} \in \mathcal{T}} y(\mathbf{s})}{|\mathcal{T}|} \right)^T.
 \end{aligned} \tag{3.5}$$

Thus, the motion estimate found by minimizing the template matching error is likely to be the motion associated with the *centroid* of the template, a result that is intuitively agreeable and is a direct extension of that for (rectangular) block matching.

As shown in Fig. 3.1, the centroid of the template \mathbf{s}_t is obviously not at the block center when the template is straddled on the top and to the left of the target block \mathcal{B} . Thus we can expect TMP to yield higher prediction error than BMC for block \mathcal{B} . A little computation using \mathbf{s}_t in place of \mathbf{s}_c in (3.3) and (3.4) further shows that the error is lower in the upper left quarter and higher in the lower right quarter. This result is well supported by the empirical data displayed in Fig. 3.2, where the actual error surface and the ones predicted by the two models are compared. For clarity we have rotated the error surfaces counterclockwise by 135° . From the figure, we also observe that Zheng's model seems to perform better in estimating error variances.

In summary, although TMP does not require extra motion information, its prediction efficiency is generally much worse than that of BMC in the mean-square error (MSE) sense. An exception is when both the intensity and motion fields are less ran-

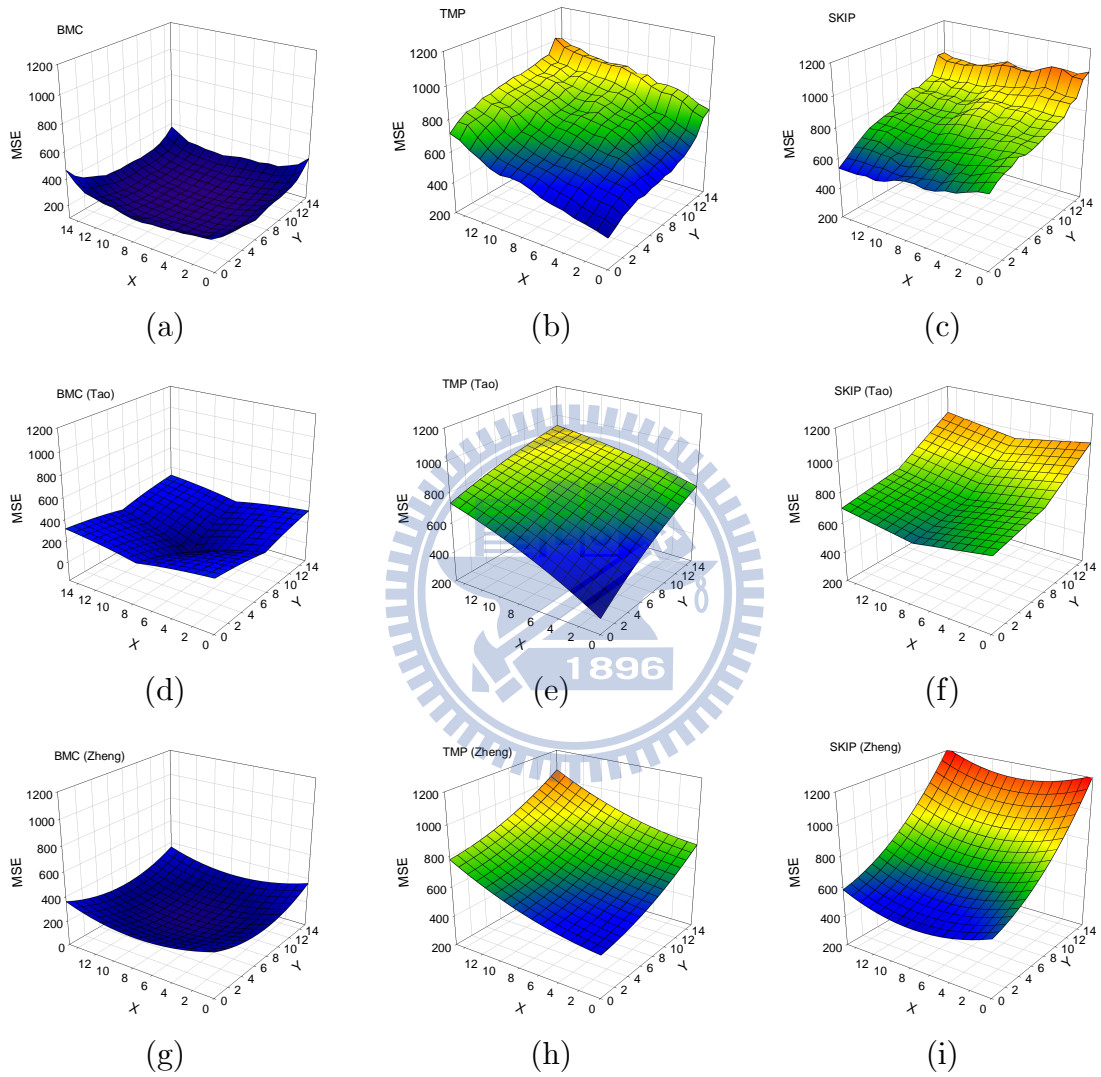


Figure 3.2: Mean-square prediction error surfaces of block \mathcal{B} produced with (a) BMC (b) TMP and (c) SKIP Prediction. The 2nd and the 3rd rows show the error surfaces predicted by Tao and Zheng’s model, respectively. The sequence is Football and the block size used for motion compensation is 16x16.

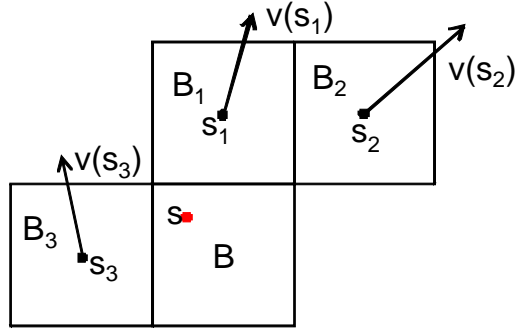


Figure 3.3: Inference scheme for SKIP Mode

dom or have high spatial correlation, that is, with Tao's model, σ_I^2, σ_m^2 are smaller or ρ_m, K tend to be larger and with Zheng's model, ϵ is small. It is then natural to question how it can achieve a bit-rate saving of 10%. The answer becomes clear when its performance is compared with that of SKIP prediction.

3.3 Error Variance Distribution of SKIP Prediction

We shall now derive formulae that will enable us to estimate the error variance for SKIP prediction. Recall that if a block is coded in SKIP mode, its motion vector is determined by the median of those in its neighborhood. Using the example shown in Fig. 3.3, the inferred vector $\hat{\mathbf{v}}$ for block \mathcal{B} is

$$\begin{aligned}\hat{v}_x &= \text{Median}\{v_x(\mathbf{s}_1), v_x(\mathbf{s}_2), v_x(\mathbf{s}_3)\} \\ \hat{v}_y &= \text{Median}\{v_y(\mathbf{s}_1), v_y(\mathbf{s}_2), v_y(\mathbf{s}_3)\}\end{aligned}\tag{3.6}$$

where $(v_x(\mathbf{s}_i), v_y(\mathbf{s}_i))^T, i = 1, 2, 3$ are the motion vectors associated with blocks \mathcal{B}_i and are approximated by the motion of their centers. The corresponding mean-square prediction error for pixel $\mathbf{s}, \mathbf{s} \in \mathcal{B}$ then becomes

$$\begin{aligned}E[d^2(\mathbf{s}; \hat{\mathbf{v}})] &= E[(I_k(\mathbf{s}) - I_{k-1}(\mathbf{s} + \hat{\mathbf{v}}))^2] \\ &= E[(I_{k-1}(\mathbf{s} + \mathbf{v}(\mathbf{s})) - I_{k-1}(\mathbf{s} + \hat{\mathbf{v}}))^2].\end{aligned}\tag{3.7}$$

Computing the expectation in (3.7), which involves *order statistics*, is in general a difficult task. To circumvent the difficulties, we take a simpler approach by assuming

that $\widehat{\mathbf{v}}(i, j) \equiv (\widehat{v}_x, \widehat{v}_y) = (v_x(\mathbf{s}_i), v_y(\mathbf{s}_j))$, $i, j = 1, 2, 3$, with each ordered pair being equally likely. Hence, we can replace (3.7) with

$$\begin{aligned} E[d^2(\mathbf{s}; \widehat{\mathbf{v}})] &= \frac{1}{9} \sum_{i=1}^3 \sum_{j=1}^3 E[(I_{k-1}(\mathbf{s} + \mathbf{v}(\mathbf{s})) - I_{k-1}(\mathbf{s} + \widehat{\mathbf{v}}(i, j)))^2], \end{aligned} \quad (3.8)$$

which can readily be evaluated by incorporating Tao's model. A straightforward calculation then gives

$$\begin{aligned} E[d^2(\mathbf{s}; \widehat{\mathbf{v}})] &= \frac{1}{9} \sum_{i=1}^3 \sum_{j=1}^3 E_M \{ 2\sigma_I^2 - 2E_{I|M}[I_{k-1}(\mathbf{s} + \mathbf{v}(\mathbf{s}))I_{k-1}(\mathbf{s} + \widehat{\mathbf{v}}(i, j))] \} \\ &= \frac{1}{9} \sum_{i=1}^3 \sum_{j=1}^3 E_M \left\{ 2\sigma_I^2 - 2\sigma_I^2 \left(1 - \frac{\|\widehat{\mathbf{v}}(i, j) - \mathbf{v}(\mathbf{s})\|_2^2}{K} \right) \right\} \\ &= \frac{2\sigma_I^2}{3K} \sum_{i=1}^3 E_M \{ (v_x(\mathbf{s}_i) - v_x(\mathbf{s}))^2 \} + \frac{2\sigma_I^2}{3K} \sum_{j=1}^3 E_M \{ (v_y(\mathbf{s}_j) - v_y(\mathbf{s}))^2 \} \\ &= \frac{4\sigma_I^2\sigma_m^2}{3K} \sum_{i=1}^3 (1 - \rho_m^{\|\mathbf{s}_i - \mathbf{s}\|_1}) + \frac{4\sigma_I^2\sigma_m^2}{3K} \sum_{j=1}^3 (1 - \rho_m^{\|\mathbf{s}_j - \mathbf{s}\|_1}) \\ &= \frac{8\sigma_I^2\sigma_m^2}{3K} \sum_{i=1}^3 (1 - \rho_m^{\|\mathbf{s} - \mathbf{s}_i\|_1}). \end{aligned} \quad (3.9)$$

Similarly, repeating the procedure in [12], we obtain the result for Zheng's model as

$$\begin{aligned} E[d^2(\mathbf{s}; \widehat{\mathbf{v}})] &\approx \frac{1}{9} \sum_{i=1}^3 \sum_{j=1}^3 E \left[\left(\frac{\partial I_{k-1}(\mathbf{s} + \mathbf{v}(\mathbf{s}))}{\partial x} (v_x(\mathbf{s}_i) - v_x(\mathbf{s})) + \frac{\partial I_{k-1}(\mathbf{s} + \mathbf{v}(\mathbf{s}))}{\partial y} (v_y(\mathbf{s}_j) - v_y(\mathbf{s})) \right)^2 \right] \\ &= \frac{1}{9} \sum_{i=1}^3 \sum_{j=1}^3 (\epsilon_x \|\mathbf{s}_i - \mathbf{s}\|_2^2 + \epsilon_y \|\mathbf{s}_j - \mathbf{s}\|_2^2) \\ &= \frac{1}{3} \sum_{i=1}^3 (\epsilon_x + \epsilon_y) \|\mathbf{s}_i - \mathbf{s}\|_2^2 \\ &= \frac{\epsilon}{3} \sum_{i=1}^3 \|\mathbf{s} - \mathbf{s}_i\|_2^2, \end{aligned} \quad (3.10)$$

where the approximation is due to the use of Taylor's expansion in computing the prediction error $I_{k-1}(\mathbf{s} + \mathbf{v}(\mathbf{s})) - I_{k-1}(\mathbf{s} + \widehat{\mathbf{v}}(i, j))$.

It is interesting to know that both (3.9) and (3.10) are merely a weighted sum of the mean-square prediction errors, i.e. $\sum_{i=1}^3 (E[d^2(\mathbf{s}; \mathbf{v}(\mathbf{s}_i))]/3)$, when $\mathbf{v}(\mathbf{s}_i), i = 1, 2, 3$

Table 3.1: Comparison of Mean-Square Prediction Error

Schemes	Football QP22			Foreman QP22			Football QP38			Foreman QP38		
	Emp.	Tao	Zheng	Emp.	Tao	Zheng	Emp.	Tao	Zheng	Emp.	Tao	Zheng
8×8 BMC	112	109	113	19	17	19	141	134	141	43	40	43
8×8 TMP L2	372	302	342	41	29	31	398	307	360	70	48	64
8×8 TMP L4	382	346	369	39	33	34	405	351	385	70	55	66
16×16 BMC	238	232	238	28	27	28	256	246	256	59	55	59
16×16 TMP L2	590	530	609	54	48	34	600	516	597	85	67	66
16×16 TMP L4	588	555	620	55	50	37	596	539	607	86	70	69
16×16 SKIP	913	916	887	129	136	140	913	914	885	329	340	339

are separately utilized for motion compensation of pixel \mathbf{s} . In fact, this is a direct consequence of our assumption made about $\hat{\mathbf{v}}$. Its validity is justified by the empirical data given in Fig. 3.2, where it is seen that the error surfaces predicted by (3.9) and (3.10) resemble closely the actual one. Also, as expected, with the help of $\mathbf{v}(\mathbf{s}_2)$ SKIP prediction tends to minimize the error at the upper part of the block, especially at the upper right quarter.

3.4 Comparison of BMC, TMP and SKIP

Table 3.1 compares the MSE of residual signals for different schemes. The empirical values and those predicted by the models are illustrated. For experiments, we use CIF Football and Foreman sequences, each being 50-frame long. The search range for block or template matching is ± 32 pixels, with quarter-pel accuracy. To simulate quantization effects, the reference frame and the template region (of size 2 or 4) are coded by H.264/AVC. In addition, the model parameters $\sigma_I^2 \sigma_m^2 / K$, ρ_m and ϵ are estimated by a least-square fit to empirical data.

From the table, several observations can be made: (a) the models are consistent with experimental results (at least qualitatively); (b) with explicit motion information, BMC yields a minimum MSE among all the schemes; (c) TMP consistently outperforms SKIP prediction regardless of the template or target block size; and (d) the MSE of TMP increases as the template or target block size is increased. The third explains why the bit rate can be significantly reduced when TMP is applied to SKIP macroblocks as an alternative prediction source [4]. The last is due to the fact that the template centroid deviates more from the center of the target block. Remarkably, these results

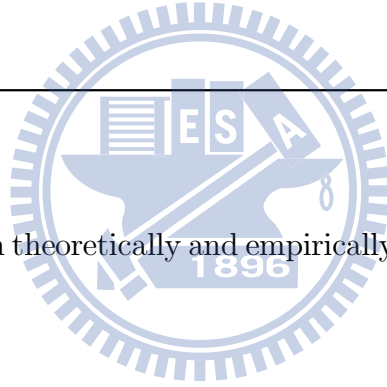
are true in an average sense, meaning that a hybrid of TMP and BMC may outperform either one alone, as reported in [4].



CHAPTER 4

Joint Application of OBMC and TMP

Having analyzed TMP both theoretically and empirically, we now look at its application to parametric OBMC.



4.1 Parametric OBMC

To reduce motion uncertainty in BMC, various forms of OBMC have been proposed in the literature. Most of their window designs, however, are constrained to use fixed block-size motion estimates. In [2], we generalize the notion of OBMC to a pixel-adaptive temporal prediction, allowing an arbitrary number of variable block-size motion estimates to be linearly combined for OBMC. Our approach there is to find a parametric solution to the optimal weights $\mathbf{w}^* = (w_1^*, w_2^*, \dots, w_L^*)$, such that the MSE for pixel \mathbf{s} , $\mathbf{s} \in \mathcal{B}$ is minimized subject to the unit-gain constraint [6]:

$$\begin{aligned} \mathbf{w}^* &= \arg \min_{\mathbf{w}} E \left\{ \left(\sum_{i=1}^L w_i d(\mathbf{s}; \mathbf{v}(\mathbf{s}_i)) \right)^2 \right\} \\ &= \arg \min_{\mathbf{w}} \left\{ \sum_{i=1}^L w_i^2 E[d^2(\mathbf{s}; \mathbf{v}(\mathbf{s}_i))] + 2 \sum_{i=1}^L \sum_{j>i}^L w_i w_j E[d(\mathbf{s}; \mathbf{v}(\mathbf{s}_i)) d(\mathbf{s}; \mathbf{v}(\mathbf{s}_j))] \right\}, \quad (4.1) \end{aligned}$$

where $\{\mathbf{v}(\mathbf{s}_i)\}_{i=1}^L$ denote the block motion vectors in some neighborhood of pixel \mathbf{s} . By adopting Zheng's model and assuming that $E[d(\mathbf{s}; \mathbf{v}(\mathbf{s}_i))d(\mathbf{s}; \mathbf{v}(\mathbf{s}_j))] = 0$ for $i \neq j$, we have, for $1 \leq i \leq L$,

$$w_i^* = \frac{r_i^{-2}(\mathbf{s})}{\sum_{i=1}^L r_i^{-2}(\mathbf{s})}, \quad (4.2)$$

where $r_i(\mathbf{s}) = \|\mathbf{s} - \mathbf{s}_i\|_2$ is the distance between \mathbf{s} and \mathbf{s}_i , the center of block i . (4.2) suggests that the optimal weight w_i^* to associate with the vector from block i should be *inversely* proportional to the *squared distance* $r_i^2(\mathbf{s})$. In [2], we also show that (4.2) performs better than the solution using Tao's model [10].

4.2 Incorporating TMP into OBMC framework

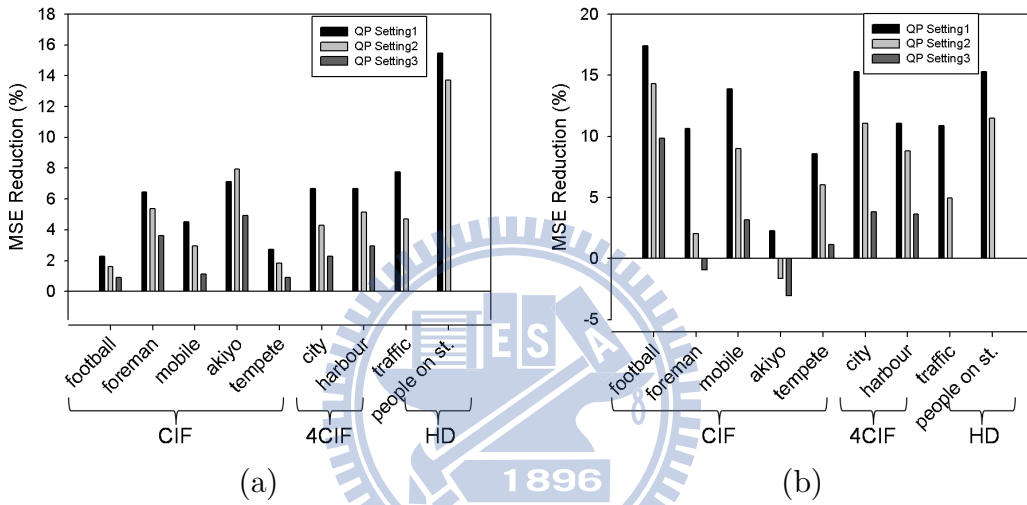
Knowing that the optimal weights are merely a function of the distances to nearby block centers, we can readily extend (4.2) to accommodate template motion estimates. A straightforward approach would consider them as the motion of template centroids and compute their weights based on (4.2). This however requires both types of motion estimates share the same model parameter ϵ , which is not likely the case. In fact, the ϵ is higher for template motion estimates. This is because with block matching, quantization noise is present only in the previous reconstructed picture, whereas with template matching, it is embedded also in the template region. The increased noise makes template motion estimates more unreliable for prediction of target blocks. Taking this into account and repeating the procedure in [2] yields *the modified optimal weights* associated with block and template motion estimates as

$$\begin{aligned} \mathbf{w}^* &= \arg \min_{\mathbf{w}} \left(\epsilon_T w_t^2 r_t^2(\mathbf{s}) + \sum_{i=1}^L \epsilon_B w_i^2 r_i^2(\mathbf{s}) \right) \\ &= \begin{cases} \frac{\widehat{r}_t^{-2}(\mathbf{s})}{\widehat{r}_t^{-2}(\mathbf{s}) + \sum_{i=1}^L r_i^{-2}(\mathbf{s})} & \text{for } w_t^* \\ \frac{r_i^{-2}(\mathbf{s})}{\widehat{r}_t^{-2}(\mathbf{s}) + \sum_{i=1}^L r_i^{-2}(\mathbf{s})} & \text{for } w_i^*, i = 1, 2, \dots, L \end{cases}, \end{aligned} \quad (4.3)$$

where the distance between \mathbf{s} and \mathbf{s}_t , the centroid of the template region, is scaled linearly, i.e., $\widehat{r}_t(\mathbf{s}) = r_t(\mathbf{s})\sqrt{\epsilon_T/\epsilon_B}$, to account for the discrepancy of model parameters.

Table 4.1: Test Conditions

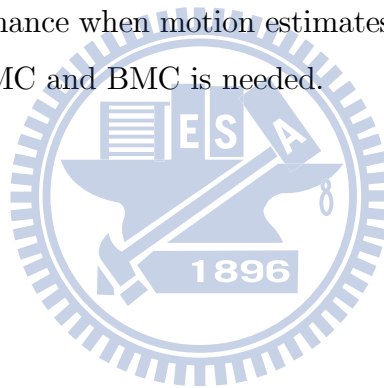
Sequences	CIF@30Hz, 4CIF@30Hz (50 frames) HD@60Hz (10 frames)
QP Settings	#1=22, #2=30, #3=38 (CIF, 4CIF); #1=22, #2=29 (HD)
Prediction Structure	1 Reference Frame + IPPP...
Search Range	CIF/4CIF/HD: $\pm 16/\pm 64/\pm 128$ pixels
Template Matching	$L = 1 \sim 4$; $N = 8$; Centroid offset: ± 10 pixels
Block Matching	Block Size: 16×16

**Figure 4.1:** MSE Reduction: (a) TMP+OBMC vs. OBMC and (b) TMP+OBMC vs. BMC.

4.3 Simulation Results

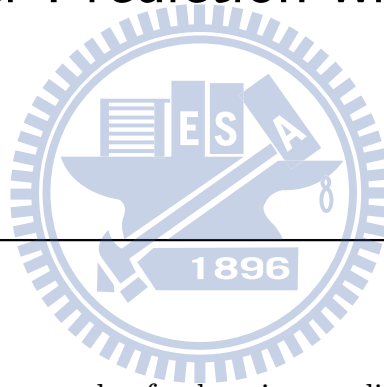
In this section, we compare the prediction efficiency of BMC, OBMC [2] and TMP when operated jointly with OBMC. The figure of merit used is mean-square prediction error. Experiments make use of video sequences in CIF, 4CIF, and HD formats, which feature a broad range of visual characteristics. To simulate quantization effects, these sequences are first encoded using H.264/AVC with different QP values. We then replicate their reference frames and template regions for motion-compensated prediction. In combining TMP with OBMC, we scale $r_t(\mathbf{s})$ by specifying a vector to relocate the template centroid and by varying the template size L (Note that the effect is more than just a linear scaling of $r_t(\mathbf{s})$). Both parameters are optimized at frame level to maximize prediction efficiency. Other test conditions are summarized in Table 4.1.

Figs 4.1 (a) and (b) show the MSE improvement when TMP and OBMC are jointly applied, with the performance of OBMC and of BMC used as a baseline, respectively. For a fair comparison, all three schemes have the same number of block motion estimates. Moreover, the search accuracy is quarter-pel. As can be seen from Fig 4.1 (a), the joint application of the two schemes can achieve a 2-16% reduction in MSE as compared with the single use of OBMC. The gain is the highest in the HD sequence "People on Street," which includes complex, random motion. This is expected as incorporating template motion estimates into OBMC can help to suppress motion uncertainty. The benefit of doing so is most obvious especially when smaller QP values are in use. This is attributed to the increased reliability of template motion estimates. Fig 4.1 (b) also confirms that the proposed method performs much better than the standard BMC except in the cases with high QP values. It was shown in Fig 4.1 that (4.2) can lead to poor performance when motion estimates are unreliable. For this, an adaptive switch between OBMC and BMC is needed.



CHAPTER 5

Enhanced Inter Prediction with TMP : P^+ Mode



In this approach, TMP plays a role of enhancing prediction efficiency when the block motion vectors are given. In last chapter, the joint application of TMP and POBMC is provided for further reduction in MSE over merely BMC. Because multi-hypotheses prediction increases memory bandwidth, it is complicated to incorporate this facility to H.264/AVC. Therefore we introduce a simplified application of TMP combined inter prediction, P^+ Mode, which performs motion estimation of current block for minimizing the weighted residual produced by TMP.

5.1 Problem Statement and Theoretical Analysis

This section mentions the encounter problems when we combine TMP and block MVs and provides theoretical solutions for these problems. As we mentioned above, P^+ Mode takes the TMP motion into consideration as estimating the motion of current block. Rather than being in place of block motion in [4] [7] [8], the motion of TMP is

applied with the current block motion for a weighted predictor. The weight function is the key for good prediction efficiency in this scheme. How does the weighting function come out? Does the optimal block MV still has higher correlation with the true motion of block center? We apply the signal models, introduced in Section 3.1, to solve these problems.

Because TMP motion is free and can be derived before block motion estimation, the contribution of TMP motion is treated as the prior information for block motion estimation. In chapter 3, the TMP tends to minimize the prediction error in the upper left quarter of a block. As a block motion is applied to minimize the residual produced by TMP, it tends to minimize the prediction error in the lower right quarter. By assuming the block MV to be found corresponds to the true motion of some unknown pixel \mathbf{b} . The problem to find the block motion can be cast as the search for that unknown pixel location \mathbf{b} that minimizes the sum of mean squared prediction error over the entire block B :

$$\mathbf{s}_b = \arg \min_{\mathbf{b}} \sum_{\mathbf{s} \in B} E [(I_k(\mathbf{s}) - w_t(\mathbf{s})I_{k-1}(\mathbf{s} + \mathbf{v}(\mathbf{s}_t)) - (1 - w_t(\mathbf{s}))I_{k-1}(\mathbf{s} + \mathbf{v}(\mathbf{b})))^2], \quad (5.1)$$

where $\mathbf{v}(\mathbf{b})$ represents the true motion of pixel \mathbf{b} and the weight factor $w_t(\mathbf{s})$ is

$$w_t(\mathbf{s}) = \frac{1 + \rho_m^{\|\mathbf{s} - \mathbf{s}_t\|_1} - \rho_m^{\|\mathbf{s} - \mathbf{s}_b\|_1} - \rho_m^{\|\mathbf{s}_t - \mathbf{s}_b\|_1}}{2(1 - \rho_m^{\|\mathbf{s}_t - \mathbf{s}_b\|_1})} \quad (5.2)$$

with model of Tao Eq.3.3 and

$$w_t(\mathbf{s}) = \frac{\|\mathbf{s} - \mathbf{b}\|_2^2}{\|\mathbf{s} - \mathbf{b}\|_2^2 + \|\mathbf{s} - \mathbf{s}_t\|_2^2} \quad (5.3)$$

model of Zheng Eq.3.4 respectively as a function of the position of \mathbf{b} when the \mathbf{b} is

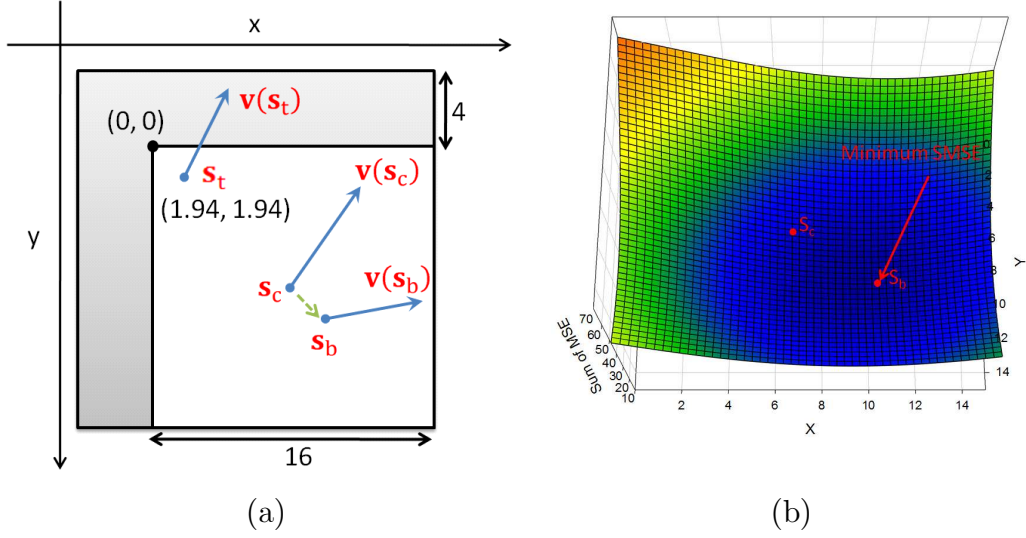


Figure 5.1: 16x16 P^+ mode: (a) an example with template centroid located at (1.94,1.94), (b) SMSE surface over different pixel \mathbf{b} , the x and y axis represent the horizontal and vertical indices of pixel \mathbf{b} .

given. Proceeding in much the same way as in [2], we obtain

$$\begin{aligned}
 \mathbf{s}_b &= \arg \min_{\mathbf{b}} \sum_{\mathbf{s} \in \mathcal{B}} E [(I_k(\mathbf{s}) - w_t(\mathbf{s})I_{k-1}(\mathbf{s} + \mathbf{v}(\mathbf{s}_t)) - (1 - w_t(\mathbf{s}))I_{k-1}(\mathbf{s} + \mathbf{v}(\mathbf{b})))^2] \\
 &= \arg \min_{\mathbf{b}} \sum_{\mathbf{s} \in \mathcal{B}} E [(w_t(\mathbf{s})(I_k(\mathbf{s}) - I_{k-1}(\mathbf{s} + \mathbf{v}(\mathbf{s}_t))) - (1 - w_t(\mathbf{s}))(I_k(\mathbf{s}) - I_{k-1}(\mathbf{s} + \mathbf{v}(\mathbf{b}))))^2] \\
 &= \arg \min_{\mathbf{b}} \sum_{\mathbf{s} \in \mathcal{B}} E [(w_t(\mathbf{s})d(\mathbf{s}; \mathbf{v}(\mathbf{s}_t)) - (1 - w_t(\mathbf{s}))d(\mathbf{s}; \mathbf{v}(\mathbf{b})))^2] \\
 &= \arg \min_{\mathbf{b}} \sum_{\mathbf{s} \in \mathcal{B}} w_t(\mathbf{s})^2 E[d^2(\mathbf{s}; \mathbf{v}(\mathbf{s}_t))] - (1 - w_t(\mathbf{s}))^2 E[d^2(\mathbf{s}; \mathbf{v}(\mathbf{b}))] \\
 &= \arg \min_{\mathbf{b}} \sum_{\mathbf{s} \in \mathcal{B}} w_t(\mathbf{s})^2 \epsilon \|\mathbf{s} - \mathbf{s}_t\|_2^2 - (1 - w_t(\mathbf{s}))^2 \epsilon \|\mathbf{s} - \mathbf{b}\|_2^2
 \end{aligned} \tag{5.4}$$

However, there is no closed form solution for the optimal pixel \mathbf{s}_b that minimizes the sum of mean squared prediction error over the entire block B . Nevertheless, an approximation of the optimal location \mathbf{s}_b can be evaluated by illustrating the relationship between sum of MSE and the location of pixel \mathbf{b} . The example for illustration is considered in Fig.5.1 (a), where the target block is a 16x16 P-MB and the template size is 4. Then Fig. 5.1 (b) plots the SMSE as a function of the location of pixel \mathbf{b} according to 5.4.

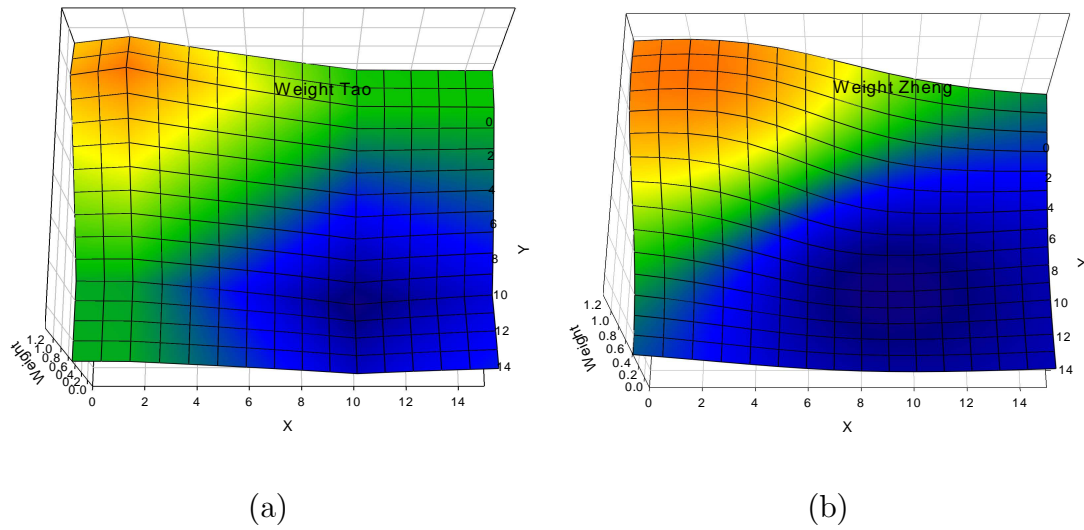


Figure 5.2: Optimal weight surfaces of P^+ Mode produced with (a) Tao (b) Zheng

From the figure, when pixel \mathbf{b} is located around (10,10), the minimum SMSE is achieved. Obviously, the optimal pixel, yielding minimum SMSE, is not the block center \mathbf{s}_c . As expected, the optimal block MV tends to be associated to the true motion of pixels in the right-lower quarter.

As the optimal location of pixel \mathbf{s}_b is given, Fig.5.2 (a) and Fig 5.2 (b) further show the window functions $w_t^*(\mathbf{s})$ of the template MV by proceeding as the approach of *optimal weights* in Eq.5.2 and Eq.5.3 respectively.

The x and y axis respectively represent the horizontal and vertical indices of a 16x16 block; the vertical axis is the weighting value of w_t for each pixel with different models. Both of the weight distributions are higher around the template centroid \mathbf{s}_t and become lower as being farther from \mathbf{s}_t . Their distributions resemble a special case of geometry partition, where the two MVs are located on the diagonal.

Fig. 5.3 provides the predicted MSEs for P^+ mode and BMC with different signal assumptions. From the figure, two observations can be made: (a) the MSE value of the pixels is lower around the motion sampling pixels i.e. \mathbf{s}_t , \mathbf{s}_c , and \mathbf{s}_b ; (b) under different assumptions of signal model, the average predicted MSEs of P^+ mode is lower than that of BMC, which is also supported in Table5.1. As our expectation, the prediction efficiency of P^+ mode, which is believed to close to bi-prediction, is better than that of BMC in a MSE sense.

In reality, the true motion of a specific pixel is hard and complicated to derived

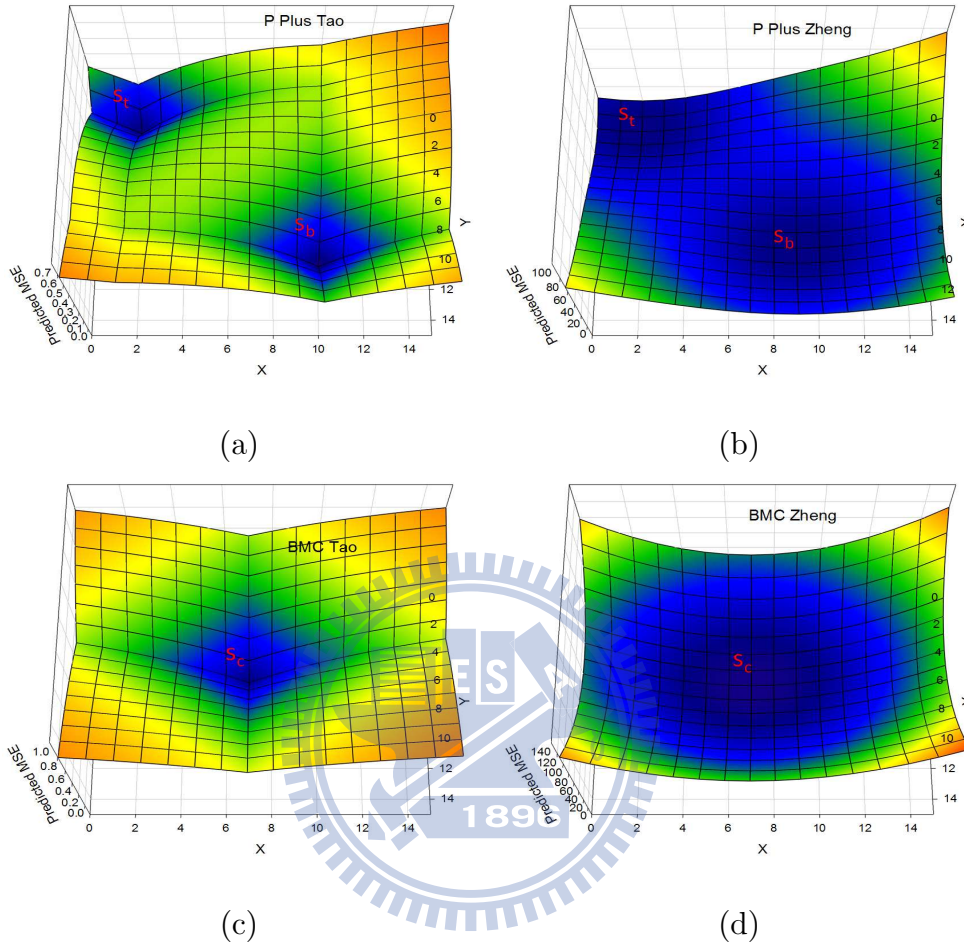


Figure 5.3: Predicted MSE surfaces of P^+ Mode produced with (a) Tao (b) Zheng and that of BMC with (c) Tao (d) Zheng

Table 5.1: Average Predicted MSE for a 16x16 block

	$P^+16 \times 16$	BMC16x16	Expected Reduction
Tao	0.366	0.542	-32.47%
Zheng	22.096	43	-48.61%

Table 5.2: Test Conditions

Sequences	1080p, 832x480, 416x240,720p
Quantization Parameter(I/P)	22/23, 27/28, 32/33, 37/38
Prediction Structure	4 Reference Frame + IPPP...
Search Range	± 128 pixels
Template Matching Range	± 32 pixels
CABAC	On

precisely, so the MV search criterion which optimizes for the new prediction mode is impractical. By noting the one-to-one relationship between $w_t^*(\mathbf{s})$ and the location of \mathbf{s}_b , the SMSE will be minimum only when $\mathbf{v}_b = \mathbf{v}(\mathbf{s}_b)$ with substitution $w_t^*(\mathbf{s})$ for $w_t(\mathbf{s})$ in 5.1. As a result, a block MV is likely to be $\mathbf{v}(\mathbf{s}_b)$ if it minimizes the sum of squared prediction error over block B:

$$\mathbf{v}^* = \arg \min_{\mathbf{v}} \sum_{\mathbf{s} \in B} (I_k(\mathbf{s}) - w_t^*(\mathbf{s})I_{k-1}(\mathbf{s} + \mathbf{v}(\mathbf{s}_t)) - (1 - w_t^*(\mathbf{s}))I_{k-1}(\mathbf{s} + \mathbf{v}))^2. \quad (5.5)$$

5.2 Experiment and discussion of P^+ Mode

In this section, P^+ Mode is merely integrated into 16x16 P-MBs in H.264/AVC reference software. The block motion search follows Eq. 5.5 for estimating the MV that minimizes the weighted MSEs as given the motion of the template. With one bit overhead for signaling to decoder for an adaptive switch between P^+ Mode and conventional BMC, P^+ Mode is incorporated into the Rate-Distortion optimization operation in H.264/AVC.

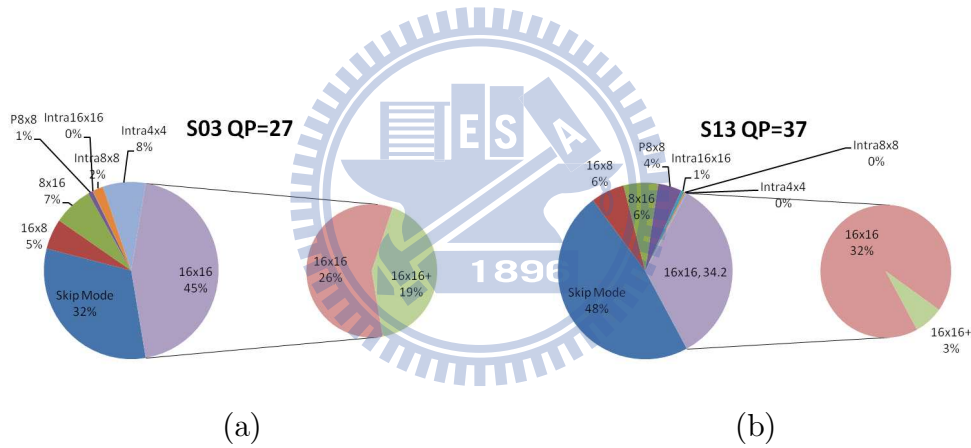
Sequences of 1080p, 832x480, 416x240 and 720p from the testing sequence set of HEVC were used and the results obtained were based on the encoding of 100 frames. Table 5.2 details the encoder settings based on H.264/AVC High Profile.

Compared to the H.264/AVC, Table 5.3 demonstrates the results in terms of BD-Bitrate savings and BDPSNR gain using the Bjontegaard tool [1]. The maximum BDBitrate saving is 4.21% (Kimono, 1080p). Average 2.42% BDBitrate reduction and 0.08 BDPSNR gain are achieved over all testing sequences.

It is observed that the gains are higher in the HD sequences and lower in the 416x240 sequences. For a deeper investigation, Fig 5.4(a) and Fig 5.4 (b) shows the mode distributions under the conditions that P^+ mode performs well and poorly respectively.

Table 5.3: Average BDBitrate Saving Over H.264/AVC

	sequences	BDBitrate	BDPSNR	
1080p	Kimono	4.21%	0.15	
	ParkScene	0.91%	0.03	
	Cactus	1.88%	0.04	
	BasketballDrive	3.12%	0.08	
	BQTerrace	3.49%	0.09	
832x480	BasketballDrill	2.64%	0.1	
	BQMall	2.20%	0.1	
	PartyScene	0.39%	0.02	
	RaceHorses	1.04%	0.05	
	BasketballPass	1.90%	0.09	
416x240	BQSquare	0.72%	0.03	
	BlowingBubbles	0.59%	0.02	
	RaceHorses	0.65%	0.03	
	720p	vidyo1	3.76%	0.14
		vidyo3	1.87%	0.07
vidyo4		2.01%	0.06	

**Figure 5.4:** Mode distribution of coding (a) S03 at QP=22, (b) S13 at QP37

From 5.4 (a), where P^+ mode performs best, it is observed that about two fifth of the 16x16 P-MBs are coded in P^+ mode. Because of the unreliability of template motion, in 5.4 (b), the low ratio of P^+ mode explains why the bitrate saving is negative of coding S13 at QP37.

Fig. 5.5 demonstrates the mode distributions of proposed scheme and anchor for encoding sequences S03 and S13 at QP=27 and QP=37 respectively. As can be seen, our scheme is shown to effective to improve coding efficiency of P16x16 mode by the observation of increased numbers of P16x16 mode at S03 QP=27. We also notice that the increased number of P16x16 MBs mostly comes from those originally coded as

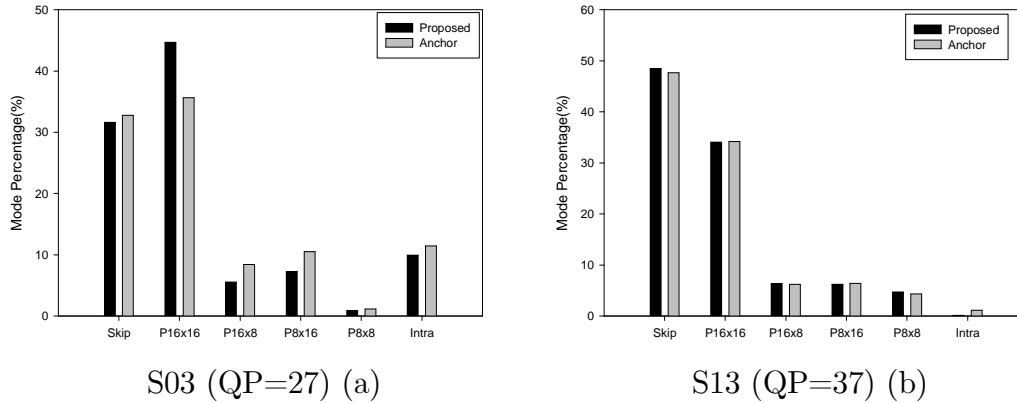


Figure 5.5: Mode distribution of proposed scheme and anchor coding (a) S03 at QP=22, (b) S13 at QP37

16x8/8x16 partition modes. This is because the proposed scheme acts as a bi-prediction mode with one free MV, it thus provides competitive coding performance over other prediction modes, especially those with two MVs such as 16x8/8x16 partition modes. On the other side, the mode distribution remain unchanged, so that one bit overhead for P^+ mode leads to a bad performance.

The detail of RD performance of proposed scheme is shown in Table 5.4 and Table 5.5.

In this section, the performance of proposed scheme are verified by directly being compared with H.264/AVC. As can be seen in the experimental result, several factors that influent the prediction efficiency of our scheme can be summarized:

- the reliability of template motion is variate with the QP parameter. It results in bad prediction performance with High QP values (equivalently, low quality settings),
- the coding mode distribution even eliminates the improvement of our scheme, and
- the higher resolution, the more improvement is. This is because the mode distribution is usually dominated by 16x16 partition.

Table 5.4: Bitrate saving and PSNR gain for each sequences and QP setting Over H.264/AVC

Resolution	sequences	QP	Bitrate Saving	PSNR Gain
1080p	Kimono	22	1.99%	0.04
		27	2.31%	0.08
		32	2.15%	0.11
		37	0.46%	0.09
	ParkScene	22	0.54%	0
		27	0.48%	0.02
		32	0.38%	0.02
		37	0.28%	0.02
	Cactus	22	1.06%	0.02
		27	1.10%	0.01
		32	1.32%	0.02
		37	1.79%	0.02
	BasketballDrive	22	2.24%	0.02
		27	2.48%	0.03
		32	1.56%	0.04
		37	1.29%	0.04
BQTerrace	22	0.68%	0.03	
	27	1.46%	0.06	
	32	1.32%	0.08	
	37	0.29%	0.08	
832x480	BasketballDrill	22	0.88%	0.01
		27	1.33%	0.04
		32	2.15%	0.04
		37	2.16%	0.07
	BQMall	22	1.06%	0.02
		27	1.39%	0.03
		32	1.47%	0.05
		37	1.34%	0.04

Table 5.5: Bitrate saving and PSNR gain for each sequences and QP setting Over H.264/AVC

Resolution	Sequences	QP	Bitrate Saving	PSNR Gain	
832x480	PartyScene	22	0.55%	-0.01	
		27	0.36%	0	
		32	0.15%	0.01	
		37	0.05%	0.02	
	RaceHorses	22	0.56%	0.01	
		27	0.68%	0.01	
		32	1.01%	0.01	
		37	0.99%	0.01	
	416x240	BasketballPass	22	0.76%	0.05
			27	0.72%	0.07
			32	0.48%	0.06
			37	0.86%	0.04
BQSquare		22	0.83%	0	
		27	0.73%	0.03	
		32	0.19%	0.01	
		37	-0.60%	0	
BlowingBubbles		22	0.45%	-0.01	
		27	0.28%	0.02	
		32	0.20%	0.02	
		37	0.08%	0	
RaceHorses	22	0.42%	0.01		
	27	0.29%	0.02		
	32	0.65%	0.01		
	37	0.74%	-0.02		
720p	vidyo1	22	1.80%	0.13	
		27	1.55%	0.12	
		32	0.35%	0.09	
		37	-0.37%	0.06	

CHAPTER 6

Conclusions

In this thesis, we have analyzed the prediction efficiency of TMP theoretically by applying two different motion and intensity models. The prediction error models of TMP are verified by empirical data and compared with those of BMC and of SKIP. From the analyses, three important observations can be made:

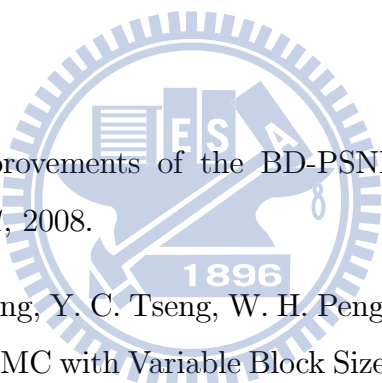
1. the motion vector of the template matching tends to be associated with the template centroid,
2. template matching can hardly compete with BMC, because of the inaccuracy of motion vector, and
3. TMP outperforms SKIP prediction, so we can explain why the bit rate can be significantly reduced when TMP is combined with SKIP prediction.

We also demonstrate in this thesis two applications of TMP. In the first application, TMP is used to enhance the BMC performance by POBMC weight function, as the block motions are already estimated. It achieves about 2-16% reductions in MSE, when compared with the single use of OBMC and an improvement of up to 18 %, as compared with the standard BMC. In the second application, an enhanced inter prediction approach, P^+ Mode, is provided to minimize the weighted residual produced

by TMP and BMC. Because block motion estimate aims to minimize the weighted residual produced by TMP, the motion does not tends to be associated with block center. The proposed scheme was integrated into 16x16 P-MBs only, but even so, we already observed an average BD-Rate saving of 2.42% and a BD-PSNR gain of 0.08dB. This work is still in its early stage. We believe there is still plenty of room for further improvement.



Bibliography

- 
- [1] G. Bjontegaard, “Improvements of the BD-PSNR Model,” *ITU-T SG16 Q.6 Document, VCEG-AI11*, 2008.
- [2] Y. W. Chen, T. W. Wang, Y. C. Tseng, W. H. Peng, and S. Y. Lee, “A Parametric Window Design for OBMC with Variable Block Size Motion Estimates,” *Proc. Int. Workshop Multimedia Signal Processing*, 2009.
- [3] T. K. T. et. al, “Intra Prediction by Template Matching,” *Proc. Int. Conf. Image Processing*, 2006.
- [4] S. Kamp, M. Evertz, and M. Wien, “Decoder Side Motion Vector Derivation for Inter Frame Video Coding,” *Proc. Int. Conf. Image Processing*, 2008.
- [5] S. Nogaki and M. Ohta, “An Overlapped Block Motion Compensation for High Quality Motion Picture Coding,” *Proc. IEEE Int. Symp. Circuits and Syst.*, 1992.
- [6] M. T. Orchard and G. J. Sullivan, “Overlapped Block Motion Compensation: An Estimation-Theoretic Approach,” *IEEE Trans. Image Processing*, vol. 3, pp. 693–699, 1994.

- [7] K. Sugimoto, M. Kobayashi, Y. Suzuki, S. Kato, and C. S. Boon, "Inter Frame Coding with Template Matching Spatio-Temporal Prediction," *Proc. Int. Conf. Image Processing*, 2004.
- [8] Y. Suzuki, C. S. Boon, and S. Kato, "Block-Based Reduced Resolution Inter Frame Coding with Template Matching Prediction," *Proc. Int. Conf. Image Processing*, 2006.
- [9] Y. Suzuki, C. S. Boon, and T. K. Tan, "Inter Frame Coding with Template Matchin Averaging," *Proc. Int. Conf. Image Processing*, 2007.
- [10] B. Tao and M. T. Orchard, "A Parametric Solution for Optimal Overlapped Block Motion Compensation ," *IEEE Trans. Image Processing*, vol. 10, pp. 341–350, 2001.
- [11] L. Wei and M. Levoy, "Fast texture synthesis using tree-structured vector quantization," *Proc. Conf. on Computer Graphics and Interactive Techniques SIGGRAPH*, 2000.
- [12] W. Zheng, Y. Shishikui, M. Naemura, Y. Kanatsugu, and S. Itoh, "Analysis of Space-Dependent Characteristics of Motion-Compensated Frame Differences Based on a Statistical Motion Distribution Model," *IEEE Trans. Image Processing*, vol. 11, pp. 377–386, 2002.